ON SEGMENTS AND SYLLABLES IN THE SOUND STRUCTURE OF LANGUAGE: CURVE-BASED APPROACHES TO PHONOLOGY AND THE AUDITORY REPRESENTATION OF SPEECH.[1]

Olivier CROUZET[2]

"There are forms within forms both up and down the scale of size. Units are nested within larger units. Things are components of other things. They would constitute a hierarchy except that this hierarchy is not categorical but full of transitions and overlaps."
James J. Gibson (1979). *The ecological approach to visual perception*.

RÉSUMÉ — SUR LES NOTIONS DE SEGMENT ET DE SYLLABE DANS LA FORME SONORE DU LANGAGE : LES COURBES EN PHONOLOGIE ET LA REPRÉSENTATION AUDITIVE DE LA PAROLE. *Les approches récentes de la syllabe réintroduisent une description continue et descriptible mathématiquement des objets sonores : les « courbes ». Les recherches psycholinguistiques sur la perception du langage parlé ont plutôt recours à des descriptions symboliques et hautement hiérarchisées de la syllabe dans le cadre desquelles segments (phones) et syllabes sont strictement différenciés. Des travaux récents sur les fondements auditifs de la perception de la parole mettent en évidence la capacité qu'ont les locuteurs à extraire une information phonétique alors même que des dégradations majeures du signal sont effectuées dans le domaine spectro-temporel. Les implications de ces observations pour la conception de la syllabe dans le champ de la perception de la parole et en phonologie sont discutées.*

MOTS CLÉS — Segments, Phonèmes, Syllabes, Perception de la parole, Représentations mentales.

SUMMARY — *Recent approaches to the syllable reintroduce continuous and mathematical descriptions of sound objects designed as "curves". Psycholinguistic research on oral language perception usually refer to symbolic and highly hierarchized approaches to the syllable which strongly differenciate segments (phones) and syllables. Recent work on the auditory bases of speech perception evidence the ability of listeners to extract phonetic information when strong degradations of the speech signal have been produced in the spectro-temporal domain. Implications of these observations for the modelling of syllables in the fields of speech perception and phonology are discussed.*

KEYWORDS — Segments, Phonemes, Syllables, Speech perception, Mental representations.

[1]The initial part of the title of this paper is freely inspired by an article that was published 20 years ago by Benoît de Cornulier: « Sur la notion de consonne et de syllabe en français » [8].

[2]Université de Nantes, Nantes Atlantique Universités, LLING - Laboratoire de Linguistique de Nantes, EA3827, UFR Lettres et Langages, Chemin de la Censive du Tertre, BP81227, Nantes, F-44000 France. Send correspondence to olivier.crouzet@univ-nantes.fr.

0. INTRODUCTION

The investigation of speech processing mechanisms in psycholinguistics has naturally led to the introduction of several concepts from phonetics and phonology into the description of cognitive language processing. Though this is still subject to intense discussions, it is generally hypothesized that representations described in linguistics may be at work during language perception and production as manifestations of a speaker's competence. Two levels of phonological description have been subject to intense investigation in the speech sciences and will be discussed here: phoneme-sized units (either of a phonetic or a phonemic nature) and syllable-sized units. These two levels of linguistic description have traditionally been viewed as rather independent aspects of phonological representations that would be involved in speech processing. Phone-sized representations are usually viewed as one of the ultimate ends of the speech perception system: the main task of speech perception is to identify the segmental content which may be associated with the corresponding acoustic signal. Syllables are usually viewed as *supra-segmental* structures that may provide (either directly or by means of their foundational phonological regularities) relevant information for both phone recognition [e.g. 15, 26, 30] and language parsing [e.g. 27, 37].

In section 1, various theoretical approaches to the syllable will be described and the associated conceptions of the relation between segments and syllables will be adressed. Section 2 will present some major issues concerning the role of phonological representations in psycholinguistics as regards syllables and will review recent work from research on the auditory bases of speech perception that provide puzzling knowledge on the perceptual organization of speech. We will then discuss the implications of these data for the theoretical understanding of the relation between segments and syllables in models of speech perception and phonology (Section 3).

1. PHONOLOGICAL APPROACHES TO THE SYLLABLE

The syllable is a major phonological entity that has proven to be fundamental in linguistics if one expects to reach an understanding of the constraints governing the sequential organisation of sounds in the languages of the world [12]. Though it is generally admitted that *syllables* are the definitive key to explaining the shaping of *sound* sequences, the question as to whether syllables are phonetic or phonological in nature has been subject to considerable debate [2, 7, 12].

Often, references to the syllable in various fields of the language sciences state (either explicitly or implicitly) that there is a *qualitative* distinction between phone-like segments and syllables. The linear chain of phones or distinctive feature matrices would therefore be organized in relation to a super-ordinate / hierarchical structure that would constitute the source for explaining universal or language specific constraints on the sequential organization of segments in human languages. (but see [6] for an alternate view in phonology).

1.1. HIERARCHICAL APPROACHES TO THE SYLLABLE

Most approaches to the syllable have described this object as a highly hierarchized structure that would be superimposed on a sequence of segments (see [12] for an extensive discussion). This hierarchy would contain various levels of description that would constitute different levels of segmental organization with respect to the phonological representation of sound sequences. The syllable's major role would be to govern the sequential distribution of segments. They would therefore exert a major influence on observations such as the general tendency for syllables in the languages of the world to restrict the number of segments that can link to each position within the syllable or the general preference for simple syllables (CV[3]) than for more complex syllables (e.g. CCCVCC)...). Phonetic realisation of segments (why some segments seem to behave differently depending on the structure of the syllable they belong to) would also be under syllabic control. In order to account for the sound structure of languages, it has been proposed that syllables are tree-like structures corresponding to various branching entities such as the onset (the first part of the syllable), the nucleus (the middle or peak of the syllable) and the coda (the final part). These objects may themselves be branches of higher structures (the rhyme associates the nucleus with the coda) and the syllable head associates all these parts together (cf. Figure 1a).

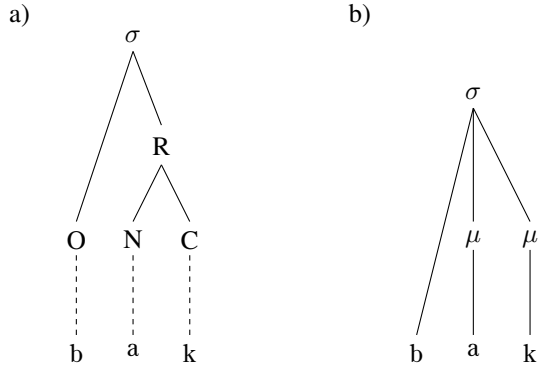

Figure 1: Syllabic representation of [bak] according to the classical onset-rhyme approach (a) and to the moraic theory (b). In a), σ stands for the syllable's head, R stands for the *rhyme*, which contains the *nucleus* (N) and the *coda* (C). The *onset* (O) directly links to the syllable's head. In b), μ stands for a *mora* weight unit.

There has been attempts at offering a *flattened* view of the syllable by means of the *moraic* theory [16, cf. Figure 1b]. According to this proposal, it is not necessary to postulate the existence of so many different *constituents* in the syllable (the onset, the nucleus, the coda and the rhyme) and limiting the organization of the syllable to a single *weight* constituent (the *mora*) should reflect fundamental properties of syllables. Though both approaches use tree-like representations, the moraic theory offers a very different understanding of what syllables *are*. The moraic theory describes the syllable as a phonological object that would be related to the weight each segment affords to the syllable. The *nature* of segments (and syllables) is therefore fundamental within this framework. According to

[3]C stands for "Consonant" and V for "Vowel".

the onset-rhyme approach the syllable is interpreted as a purely hierarchical and symbolic structure. Such a hierarchical framework endorses a view that is typical of the *symbolic* paradigm in cognitive science [29, 36]: the syllable as it is described depends on the usage of complex entities (onsets, nuclei, codas, rhymes) and representations (hierarchical trees) that would result from the application of symbolic rules or constraints inherited from classical rule-based approaches to cognition (see [23] for a general discussion of this issue). It may however be argued that purely hierarchical approaches have mostly described constraints for identifying syllable *peaks* and *boundaries* or for applying syllabification algorithms but have not provided any *explanation* as to what a syllable *is* (for an extensive discussion on this matter, see [2]).

Within the hierarchical framework, the nature of segments (their *substance*) has been refered to in order to maximize the predictions of syllabification algorithms and / or to account for various patterns that may not be accounted for in terms of external rules (e.g. why is [bʁa] –eng. *arm*– a possible word in french but not *[ʁba]?). However, their use of substance is simply a means to construct a hierarchical representation of the syllable. In the line of proposals that had been made by the past [17], recent approaches have tried to suppress reference to such *constituents* of the syllable (onset, nucleus, coda...), involving only non-symbolic (or sub-symbolic) representations[4].

1.2. Substance-based and sub-symbolic approaches to the syllable

Such alternative approaches to hierarchical and symbolic views of the syllable, though rarely refered to in psycholinguistic research, are certainly among the most influential frameworks in linguistics today. When they are advocated in the field of psycholinguistics, they are more often viewed through the prism of an opposition between linguistic vs. probabilistic (or phonetic) mechanisms rather than as linguistic representations proper. According to these approaches, syllables and their constituents simply *emerge* from very basic linguistic mechanisms and representations. It is often argued that taking into account the nature of segments and describing this nature as the evolution of a parameter on a continuous dimension should offer sufficient power for explaining syllabic objects and sound structures in the world's languages. These characteristics may either be articulatory [5] or both articulatory and auditory [20] but may also include more abstract mathematical objects [1, 2]. Approaches outside the articulatory phonology framework are usually based on one of the most popular substance-based hypotheses in the phonological study of syllables: the *sonority scale* (see [7] for a review). In phonology, proposals refering to variables that may vary on a continuous dimension are identified as *curve-based* models.

[4]The term *non-symbolic* representations is ususally refered to in cognitive science to identify objects that may not be described with linguistic formulations. It is obvious that some of the non-symbolic approaches wich will be described in the next part of this article would be described as perfectly *symbolic* according to mathematics (a sinusoïdal function *is* a symbolic object). By *non-symbolic*, therefore, we mean objects that may be described in terms of mathematical functions while *symbolic* models usually refer to approaches centering on complex representations that may no be easily described with mathematical functions (e.g. "*a coda* is the last part of the syllable").

Laks [21] shows that *curve-based* models exhibit properties that are similar to those reflected by the use of *constituents* such as onsets, nuclei, codas... Syllabic constituency properties may then be conceived as emerging from the interactions between more fundamental parameters of the representation of phonological forms. For the sake of computational simplicity, Laks argues that there is no need to use constituents such as those described in hierarchical / symbolic models.

1.2.1. Syllables in "Articulatory Phonology"

Though phonology has mostly employed articulatory descriptions when analyzing the nature of segments by means of distinctive features, *Articulatory Phonology* rejects the usual conception of phonological features as sequences of discrete matrices. According to this framework [4, 5], the sound structure of language is built on the organization of simple *speech gestures* (tongue elevation, vocal cord vibration, lip closure...) that are dynamical in nature. Gestural specifications are defined continuously in time and may overlap. Syllables are therefore viewed as a combination of gestural events that extend over time, exhibit temporal variation and may overlap at different degrees. Phonological gestures are usually represented by means of several continuous curves, each representing the continuous evolution of a single gesture in time. Articulatory phonology does not make reference to acoustic properties of speech signals. Though this statement may sound rather weird, the articulatory phonology framework may potentially extend to acoustic properties of speech. Even if the general framework according to which it has been developed is related to an understanding of speech as "*[physiological] movements made audible*", its fundamental principles lie on a *dynamical* view of phonological organization: phonological representations are described as simple patterns of coordinated organization. The question as to whether there would be a need or an opportunity for such an evolution will be addressed in section 3 (nevertheless, see [31, 35] for related approaches).

1.2.2. The sonority scale

Outside the Articulatory Phonology framework, most current approaches to the syllable make extensive use of the *sonority scale* in their definition of the syllable. Historically, this scale has first been described in favor of a *phonetic* interpretation of syllables. However, physical correlates of this multilevel scale have not been fully identified though coarse phonetic principles may apply [7]. Basically, sounds are described as varying in their degree of sonority on a continuous scale, going from stops (the least sonorous sounds) to open vowels (the most sonorous sounds). Several physical principles underlying this scale have been discussed among which the openness of the vocal tract (on the physiological side) and the corresponding amount of acoustic energy (on the acoustical side). Syllabification of segments, e.g. the organisation of segments into syllables, is described as the implementation of constraints on the evolution of the sonority curve within the syllable. It is argued that a well-formed syllable may only contain a sequence of segments for which sonority rises monotonously from the first segment to the peak then falls down monotonously again until the last segment within the syllable (cf. Fig-

ure 2). Klein [20] combines this sonority scale with a *consonanticity* scale and parallels this proposal with a corresponding mirroring of articulatory and perceptual *tiers* (cf. next section). According to Clements [7], nothing may currently lead to accept the idea of a sonority scale as an actual physical parameter of speech production (neither in the articulatory nor in the acoustic domain). However, it may be argued that this scale reflects properties that are attributed to the *abstract* phonological representation of these sounds.

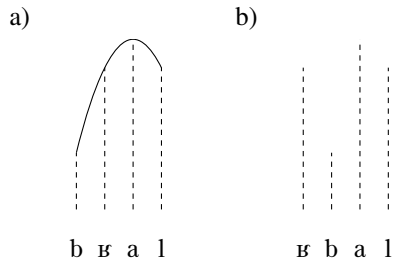

Figure 2: Illustration of the constraints on the sonority curve for the syllables [bʁal] and *[ʁbal]. The latest one (b) would not be considered as well-formed due to a decrease in sonority between the first and second consonants. But see [22] for a more thourough approach.

Of course, some languages may actually accept sequences that would not be well-formed in other languages. There are still debates on universals and language specific constraints concerning syllabic models and their capacity to account for the various observations in the languages of the world. Exceptions to the predictions derived from such basic models have led to developing various solutions without abandonning the fundamental principles. Laks [22] shows that *relative* well-formed sonority patterns may emerge from the output of a connectionnist network, which accounts for french syllabification better than *absolute* sonority values: each segment may influence its neighbours' intrisic sonority level. Another approach to the limits afforded by the sonority scale alone has been to introduce the notion of rhythmicity or cyclicity in syllabic organisation.

1.2.3. Syllables as rhythmic / cyclic structures

This approach to syllables as objects that would be determined by the regular modulation of sonority within the syllable has led to rhythmic / cyclic descriptions in syllabic theory. If sonority is cyclic *within* the syllable, it should therefore cycle *between* syllables. According to such a statement, syllables may be described as cyclic objects. Angoujard [2] describes syllables as the association of a sonority (or prosodic) curve with a *rhythmic grid* (cf. Figure 3). The syllable is the result of the relation between the sound substance (the sonority scale) and the syllabic rhythm (the rhythmic grid). Klein [20] does not make use of separate sonority and rhythmic curves but uses a double-sided substance curve which is based on both sonority and consonanticity (cf. Figure 4). According to both views, syllables are described as the output of a mechanism which links *substantial* segments with one or several cycling alternations (the rhythmic grid and / or the sonority – consonanticity curve).

The notion of cyclicity in speech and language has been addressed in many areas of speech science. This is congruent with hypotheses concerning the emergence of speech in

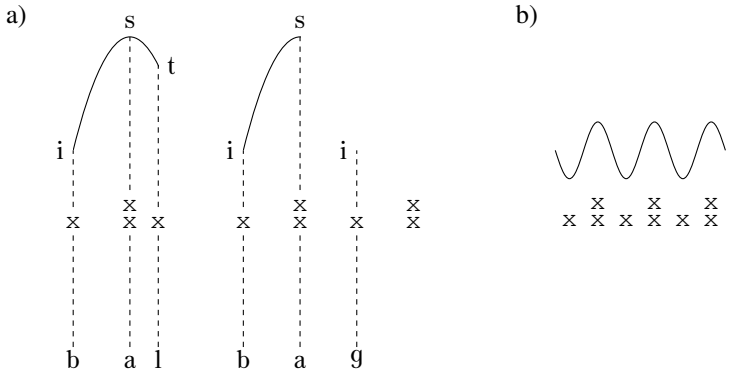

Figure 3: a) Syllabic representation of [bal] and [bag] according to Angoujard [2]. The bottom line contains the sequence of segments; on the middle line, segments link to the *rythmic grid* that represents rythmic alternations between strong and weak positions; the top line (the curve) is described as a representation of the segmental sonority scale. The bottom line segments are usually represented as segmental units but are here for simplification and represent matrices of *elements* [19]. b) The rhythmic grid may be analyzed as a continuous –sinusoïdal– modulation.

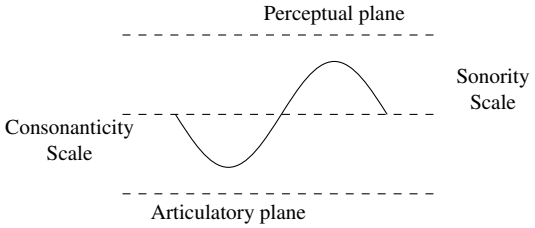

Figure 4: The syllable according to Klein [20]. The syllable is a rhythmic modulation of sonority / consonanticity to which segments may link to. Where they link on this curve may influence their realization. Consonants tend to link to the left part of the curve (under the middle straight line) and vowels tend to link to the right part of the curve (over the middle straight line). If a vowel links far from the sonority peak, its sonority will decrease and it may be pronounced as a glide.

newborns [25] as well as with approaches to speech production [32, 35]. Such proposals are largely developed within the cognitive sciences by means of the *dynamical systems* framework [10, 33]. Concerning the relation between segments and syllables, it seems obvious that segments and syllables are not independent from one another. The nature of segments determines which syllables may be produced in a given language. In return, the nature of segments may be influenced by their position on these cyclic curves. Still, all these proposals may have in common that segments be described as either *discrete* or *abstract* with respect to time. According to Angoujard [2], the rhythmic grid is represented as a discrete sequence of rhythmic events. However, such a discrete representation of a rhythmic alternation may be reduced to a continuous modulation of a rhythmic parameter (cf. Figure 3b). According to [20], no specific considerations are offered concerning this issue. However, Brandão de Carvalho [6], refering to this work, argues in favour of a description of phonological features as spanning over several "segments" within the syllable.

2. SPEECH PROCESSING, SEGMENTS AND SYLLABLES

Since the advents of *Generative Grammar*, linguistic structures have been described as constituents of a speaker's *competence*. As such, they are hypothesized as *mental representations* that have to be developed or discovered during language acquisition. The question as to whether these representations are at work during speech perception and production is still debated and will only partly be adressed in this section. The involvement of syllabic representations in speech perception has been investigated in the fields of phonetic identification and lexical parsing. Hierarchical descriptions of the syllable have been favoured in these studies. Yet, recent research on the auditory bases of speech perception contribute to reinterpreting the analysis of the segment / syllable distinction.

2.1. SYLLABIC EFFECTS IN PSYCHOLINGUISTICS

The influence of phonological regularities on speech identification processes has been evidenced using various behavioral tasks. It has been observed that phonetic identification seems to be influenced by *phonotactic regularities* in american english speakers [26]. When a synthetic ambiguous consonant between [r] and [l] is placed within a CCV sequence, the identification of the median ambiguous segment seems to depend on the phonotactic acceptability of the consonant cluster that would be formed by the initial sequence of consonants. When the ambiguous sound is preceded by a [s], the ambiguous segments tends to be identified as a [l] rather than as a [r]; when it is preceded by a [t], the ambiguous segments tends to be identified as a [r] rather than as a [l]. This tendency is exhibited in a lateral displacement of the *categorical boundary*. Similar results have been observed with french speakers [15]. Though several interpretations of this effect may compete [30], phonotactic constraints are viewed as higher level regularities that may govern the interpretation a listener gives of a physical segment.

Syllabic effects have also been observed in word processing mechanisms [27, 37]. When participants are asked to detect real monosyllabic words embedded at the beginning or at the end of nonsense sequences (*word-spotting*), listeners experience more difficulty detecting the word when the boundary between the lexical and the non-lexical parts do not coincide with the syllabic boundary; detection is easier when these boundaries are congruent. For example, Dutch speakers experience more difficulty to detect 'rɔk' –*skirt*– in /fim.rɑk/ (where the '.' indicates a syllabic boundary) than in /fi.drɑk/ [27]. Vroomen & de Gelder [37] investigated a similar issue with a phoneme monitoring task in which speakers had to monitor target phonemes inside sentences. The target was either pronounced at the coda (as in 'de.boo**t**.die.ge.zon.ken [...]'; in these examples the target is a [t] and appears in bold font) or at the onset (in 'de.boo.**t**is.ge.zon.ken [...]') of the syllable which followed the monosyllabic word. In the onset condition, there was a misalignment between the syllabic and lexical boundaries; these were aligned in the coda condition. Longer reaction times were observed when the target phoneme was in onset position than in the coda condition. Though diphone frequencies may account for these results, the interpretation of the relation between segments and syllables is also interpreted in terms of a *hierarchical* influence of syllabic structure. Though these observations seem

to suggest that the representation of syllabic structure is involved in psycholinguistic processes such as phonetic identification or word recognition, it is still possible that these effects may be accounted for by non-syllabic influences like sequential diphone frequencies or lexical competitions among others.

On an alternate viewpoint, the nature of the relation between a phonetic and a syllabic level of representation has to be adressed. As a matter of fact, if one states that segment recognition or lexical segmentation refer to –and may be influenced by– syllabic structure, it is often implied that there *are* two different levels of representation involved in the cognitive processing of speech that may account for these data: a segmental level involving phone-sized segments and a supra-segmental level representing their structured organization.

2.2. AUDITORY PERCEPTION OF PHONE-SIZED SEGMENTS

Though research in psycholinguistics has taken hierarchical approaches to the syllable as a major reference of the relation between segments and syllables, it is not clear whether all approaches to the cognitive processing of speech by listeners would favor such an hypothesis. Indeed, current research on the auditory bases of speech perception in quiet or in degraded environments provides cues to questionning this distinction between various levels of representation in linguistics.

2.2.1. On the time-span of spectral content

When 4 sinewave tones are presented sequentially to human listeners, correct judgements of order only occur for individual durations larger than 200 ms. At shorter durations, listeners may still discriminate between groups of tones (or noises) but may not tell the order in which individual components occurred within the sequence. It is inferred that listeners exhibit perception of the global pattern of sounds but may not analyse the composition of this *compound*. It is often admitted that in spontaneous speech, mean segment durations are about 70–80 ms. One of the hypotheses that have been offered for explaining the astonishing aptitude of speakers to decipher the ordering of e.g. CCV sequences has been to predict that thresholds for order identification would be much lower for speech than for non-speech. However, when tones are replaced with steady-state (synthetic) vowels ranging in individual duration from 30 to 100 ms, the component vowels may not be identified and *verbal temporal compounds* are perceived: listeners report hearing words corresponding to sequences of consonants and vowels. Nevertheless, when these vowels are presented in isolation, they may still be identified correctly (see [38] for a review).

Though vowel duration seems to impose restrictions on how speech signals may be interpreted as sequences of sounds, one may state that in natural speech there is no need for deciphering the order of component segments as, phonetically, signals may be described as containing overlaping acoustic information [24]. However, Saberi & Perrott [34] report that when all consecutive segments of a speech signal are temporally *reversed*, it may still be possible to extract phonetic information. Saberi & Perrott proceeded to the inversion of temporal slices of speech signals for windows ranging from 10 ms to 300 ms.

When participants listened to locally time-reversed sentences with all segments reversed in time, perfect intelligibility occurred up to 50 ms reversal windows. Proportions of approximately 70-75% intelligibility were observed for 100 ms segments and still 50 % intelligibility was reported with a window length of 130 ms. As a matter of fact, performance started to decrease severely for temporal windows corresponding to the *ultralow* frequency modulation of speech envelopes (c.a 3 to 8 Hz). These *frequency modulation* components correspond to durations ranging from approximately 125 ms (8 Hz) to 333 ms (3 Hz). Obviously, these are durations that are typical of *syllable-sized* segments [14]. It seems that, as far as syllable-size information is preserved, *phonetic information may still be extracted from the acoustic signal*. Note however that, according to Greenberg & Arai [13], the authors asked participants to *rate* the intelligibility of sentences and did not actually estimated this intelligibility level. In the next section, attested intelligibility levels are presented from Greenberg & Arai's work [13] that correspond only approximately to the pattern reported by Saberi & Perrott but that will still favour a specific view of the relation between segments and syllables.

2.2.2. Spectro-temporal perceptual organization

When locally time-reversed signals are presented, global spectral information –though reversed for these portions of time– is still roughly present within these time-scales. Even though limits on intelligibility seem to correspond to syllabic durations, one may argue that reversed spectral information still provides sufficient data for phonetic identification and that only larger amounts of reversal prevent phonetic identification from occurring. Such an interpretation may follow from Warren's work on the perception of short vowel sequences and their identification as temporal compounds. Phonetic perception may depend on a global spectral pattern and would resist to reversal up to certain limits. This *ultralow* (3 to 8 Hz) modulation limit may just happen to coincide with syllable-size durations. What happens then when spectral content does not appear at the same time in speech signals ?

Arai & Greenberg [3] applied desynchronization of narrow spectral bands to sentence recordings. According to the authors, this procedure produces alterations of the speech signals that are partly similar to those produced by *reverberation*. Each signal was split into 19 frequency channels which were pseudo-randomly and uniformly delayed in time with individual delays ranging from 0 ms to D_{max} (where D_{max} ranges from 60 to 240 ms). For a given signal, the mean delay was equal to $D_{max}/2$. Listeners reported almost perfect word recognition scores up to a maximum temporal delay (D_{max}) of 140 ms, where they still reached 75% word recognition performance. The authors note that even when the asynchrony exceeded 200 ms, performance was still as high as 50 % correct word recognition. Again, almost perfect phonetic identification occurs for temporal degradation up to 140 ms, which corresponds to 70 ms average desynchronization of spectral channels in time. According to these data, it seems possible to extract phonetic information even when two or more different phonetic segments overlap in time with respect to their spectral content.

In a subsequent study, Greenberg & Arai [13] replicated Saberi & Perrott's experiment in order to investigate the relationship between the observed data and a specific model of the auditory bases of speech perception as well as to register "actual" intelligibility performance rather than intelligibility "estimations" from listeners. Time-windows were locally reversed for durations ranging from 0 to 180 ms by steps of 20 ms. Intelligibility performance for word recognition was close to perfect up to 40 ms window-length (80% correct word recognition). Intelligibility then decreased abruptly to 25% for 80 ms window-length and reached an asymptote (4% correct recognition) at 100 ms. Though Greenberg & Arai's results seem to contradict Saberi & Perrott's data with respect to the relationship between reversed-segments' duration and speech intelligibility, Greenberg & Arai observe that intelligibility performance is highly correlated with both the phase and amplitude of the *modulation spectrum* (see [18] for an overview), a measure of the low modulation components of a single spectral channel and of the phase angle between the original channel and its modified (locally reversed) version. It appears that speech intelligibility declines correlatively to the magnitude of the 3 to 8 Hz components of this complex modulation spectrum. As was hypothesized in the previous section, it seems that speech recognition processes are dependent upon the integrity of this 3 to 8Hz modulation component which broadly corresponds to syllable durations in spontaneous speech. As far as this portion of the modulation spectrum is preserved, perfect phonetic identification may occur. Only when it is degraded would phone recognition vanish.

3. ON REPRESENTATIONS

3.1. ON THE REPRESENTATION OF SEGMENTS AND SYLLABLES

In this paper, current theoretical approaches to segments and syllables have been described. If two conceptions of the syllable compete in phonology (constituent-based vs. curve-based approaches), general reference to the syllable in language processing regularly refer to syllables as highly hierarchized structures organized into constituents. In tree-like views of the syllable, linear chains of segments are structured according to a "governing" syllabic organization. Curve-based approaches generally define the syllable as the emergent property of several simple representations which mostly involve objects that may be described in continuous mathematical terms. Yet, *segments* are still mostly represented as *discrete* events or will sometimes link to *discrete* squeletal positions. Nevertheless, the discrete nature of segments may itself be an emergent property of phonological organization and some theoretical accounts provide bases for such a proposal [2, 6, 20].

According to psycholinguistic research concerning the influence of syllabic structure on phonetic identification and lexical parsing, syllabic organization seems to govern the *interpretation* of segments and the *localisation* of lexical boundaries. However, research on the auditory bases of speech perception seems to contradict a view of the syllable as an ordering structure that would *contain* discrete segments: segmental information may be afforded by *global* temporal modulations. Fine phonetic details need not

be present for phonetic identification and it is the phase and amplitude components of the modulation spectrum within the syllable-sized temporal domain that seem to contain most of the relevant information for speech processing. This modulation spectrum may be viewed as a *local* temporal modulation curve that reflects properties of speech sounds for several segments at once. According to Arai & Greenberg [3], "*such intelligibility data are difficult to reconcile with spectral models of speech recognition*".

3.2. On the relationship between mental representations and phonological objects

Linguistic representations are usually described as part of a speaker's competence within traditional generative linguistics. Linguistic "knowledge" is consequently supposed to occur during the processing of speech signals. Though various experimental observations have contributed to the temporary conclusion that phonological constraints may influence the processing of speech signals, data concerning the auditory bases of speech perception seem to dismiss a view of speech processing that would involve different levels of representation roughly corresponding to phone-sized and syllable-sized domains. Though these research need further investigations in order to specify the exact kind of phonetic information that may be afforded by such temporal modulations, they lead to question the status of phones and syllables in speech perception and, in return, in phonology. Two approaches to syllable representation compete though it has been shown that constituency-based approaches may simply be viewed as emergent properties of curve-based proposals. On the speech processing side, effects on phonetic identification and lexical parsing have been interpreted in terms of the structural influence of syllabic organization on the identification of speech segments and on the formation of word-boundary hypotheses. Yet, the interpretation of data from research on the auditory bases of speech recognition seems to favour a view of the syllable that would not "govern" the organization of segments but that would litterally "be" these segments...

As a matter of fact, similar issues have been adressed by the past concerning speech processing mechanisms [8, 28] though the corresponding interpretations have finally been ruled out in favour of hierarchical analyses of syllabic effects. We hypothesize that parts of the explanation lie in classical cognitivist approaches to mental representations and processes. According to such approaches, each mental "state" exists in the speakers' mind. Either phonemes *or* syllables or both of them are mental representations but they must exist in the mind of a speaker under any relevant form; and, *if they exist simultaneously*, they must do so as independently as possible –at most they may interact but should not constitute *overlapping* properties of the same physical signal. What the dynamicist approach to cognitive science [10, 33] offers is a way to view the various entities that have been described in phonology as properties emerging from the *codetermination* of fundamental sensori-motor and mental /phonological mechanisms [9, 11]: several levels of temporal resolution may then be observed simultaneously without ever being part of a speakers' mental states [29, 36]: only fundamental modulations would be at work during speech processing. These modulations may certainly refer to both articulatory, acoustic

and mental descriptions of speech signals.

We argue that the data reviewed in this article contribute to favouring a curve-based approach to syllables in both speech perception and phonology. Segments and syllables are not qualitatively distinct from one another, they just represent two different levels of resolution (among others) that may dynamically emerge from a representation of the sound structure of language in terms of simple continuous functions. This view seems congruent with recent phonological proposals concerning the organization of features within syllables [6]. These proposals have been developped in view of current knowledge on the auditory bases of speech perception and will need further investigation in order to evaluate the amount and type of featural information that is provided by the modulation spectrum for segment recognition.

REFERENCES

[1] Angoujard, J.-P., "Pourquoi des courbes ?", In B. Laks & M. Plénat (Eds.), *De natura sonorum : essais de phonologie*, Saint-Denis, France: Presses Universitaires de Vincennes, 1993.

[2] Angoujard, J.-P., *Théorie de la syllabe*. Paris, France: CNRS Éditions, 1997, 224 p.

[3] Arai, T., & Greenberg, S., "Speech intelligibility in the presence of cross-channel spectral asynchrony", In *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, pp. 933–936, Seattle, USA, 1998.

[4] Browman, C. P., & Goldstein, L., "Some notes on syllable structure in articulatory phonology", *Phonetica*, *45*(2–4), 140–155, 1988.

[5] Browman, C. P., & Goldstein, L., "Articulatory phonology: an overview.", *Phonetica*, *49*(3–4), 155–180, 1992.

[6] Brandão de Carvalho, J., "What are phonological syllables made of? The Voice / Length symmetry", In J. Durand & B. Laks (Eds.), *Phonetics, Phonology and Cognition*, pp. 51–79, Oxford, UK: Oxford University Press, 2001.

[7] Clements, G. N., "The role of the sonority cycle in core syllabification", In J. Kingston & M. E. Beckman (Eds.), *Papers in laboratory phonology I: Between the grammar and physics of speech*, pp. 283–333, Cambridge, UK: Cambridge University Press, 1990.

[8] de Cornulier, B., "Sur la notion de consonne et de syllabe en français", *Linguisticæ Investigationes*, *10*, 275–287, 1986.

[9] Fowler, C. A., "An event approach to the study of speech perception from a direct-realist perspective", *Journal of Phonetics*, *14*(1), 3–28, 1986.

[10] Gafos, A. I., & Benus, S., "Dynamics of Phonological Cognition", *Cognitive Science*, *30*, 1–39, 2006.

[11] Gibson, J. J., *The Ecological Approach to Visual Perception*. Hillsdale, NJ, USA: Lawrence Erlbaum Associates, 1979, 1986.

[12] Goldsmith, J. A., *Autosegmental and Metrical Phonology*. Cambridge, Mass.: Basil Blackwell, 1990.

[13] Greenberg, S., & Arai, T., "The relation between speech intelligibility and the Complex Modulation Spectrum", In *Proceedings of Eurospeech'2001*, Aalborg, Denmark, 2001.

[14] Greenberg, S., Carvey, H., Hitchcock, L., & Chang, S., "Temporal Properties of Spontaneous Speech – A syllable-centric perspective", *Journal of Phonetics*, *31*, 465–485, 2003.

[15] Hallé, P., Seguí, J., Frauenfelder, U., & Meunier, C., "The processing of illegal consonant clusters: A case of perceptual assimilation?", *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 592–608, 1998.

[16] Hyman, L., *A theory of phonological weight*. Dordrecht: Foris Publications, 1985.

[17] Jespersen, O., *Lehrbuch der Phonetik*. Leipzig & Berlin, GE: B. G. Teubner, 1904.

[18] Kanedera, N., Arai, T., Hermansky, H., & Pavel, M., "On the relative importance of various components of the modulation spectrum for automatic speech recognition", *Speech Communication*, *28*, 43–55, 1999.

[19] Kaye, J., Lowenstamm, J., & Vergnaud, J.-R., "The internal structure of phonological elements: A theory of charm and government", *Phonology*, *7*(2), 193–231, 1990.

[20] Klein, M., "La syllabe comme interface de la production et de la réception phoniques", In B. Laks & M. Plénat (Eds.), *De natura sonorum : essais de phonologie*, pp. 101–142, Saint-Denis, France: Presses Universitaires de Vincennes, 1993.

[21] Laks, B., "La constituance revisitée", In B. Laks & M. Plénat (Eds.), *De natura sonorum : essais de phonologie*, pp. 173–220, Saint-Denis, France: Presses Universitaires de Vincennes, 1993.

[22] Laks, B., "A connectionist account of French syllabification", *Lingua*, *95*, 51–76, 1995.

[23] Laks, B., *Langage et Cognition : L'approche connexionniste*. Paris: Lavoisier, 1996.

[24] Liberman, A., Cooper, F., Shankweiler, D., & Studdert-Kennedy, M., "Perception of the Speech Code", *Psychological Review*, *74*(6), 431–461, 1967.

[25] Mac Neilage, P. F., "The frame / content theory of evolution of speech production", *Behavioral & Brain Sciences*, *21*, 499–546, 1998.

[26] Massaro, D. W., & Cohen, M. M., "Phonological context in speech perception", *Perception & psychophysics*, *34*(4), 338–348, 1983.

[27] McQueen, J. M., "Segmentation of continuous speech using phonotactics", *Journal of Memory and Language*, *39*, 21–46, 1998.

[28] Mehler, J., Dommergues, J. Y., Frauenfelder, U., & Segui, J., "The syllable's role in speech segmentation", *Journal of Verbal Language and Verbal Behavior*, *20*, 298–305, 1981.

[29] Petitot, J., Varela, F. J., Pachoud, B., & Roy, J.-M., "Combler le déficit : Introduction à la naturalisation de la phénoménologie", In J. Petitot, F. J. Varela, B. Pachoud, & J.-M. Roy (Eds.), *Naturaliser la phénoménologie. Essais sur la phénoménologie contemporaine et les sciences cognitives*, pp. 1–100, Paris: CNRS Éditions, 2005.

[30] Pitt, M. A., "Phonological processes and the perception of phonotactically illegal consonant clusters", *Perception & psychophysics*, *60*(6), 941–951, 1998.

[31] Plaut, D. C., & Kello, C. T., "The emergence of phonology from the interplay of speech comprehension and production: A distributed connectionist approach", In B. MacWhinney (Ed.), *The emergence of language*, pp. 381–415, Mahwah, NJ: Lawrence Erlbaum, 1999.

[32] Port, R. F., "Meter and Speech", *Journal of Phonetics*, *31*, 599–611, 2003.

[33] Port, R. F., & van Gelder, T. (Eds.), *Mind as motion: Explorations in the dynamics of cognition*. Cambridge, MA: Bradford books / MIT Press, 1995.

[34] Saberi, K., & Perrott, D. R., "Cognitive restoration of reversed speech", *Nature*, *398*, 760, 1999.

[35] Tuller, B., & Kelso, J. A., "The production and perception of syllable structure", *Journal of Speech and Hearing Research*, *34*(3), 501–508, 1991.

[36] Varela, F., *Invitation aux Sciences Cognitives*. Paris: Seuil, 2001.

[37] Vroomen, J., & de Gelder, B., "Lexical access of resyllabified words: Evidence from phoneme monitoring", *Memory and Cognition*, *27*(3), 413–421, 1999.

[38] Warren, R. M., "The Relation of Speech Perception to the Perception of Nonverbal Auditory Patterns", In S. Greenberg & W. A. Ainsworth (Eds.), *Listening to Speech: An Auditory Perspective*, pp. 333–349, Mahwah, NJ, USA: Lawrence Erlbaum Associates, 2006.